

## State space models with spatial deformation

Fidel Ernesto Castro Morales · Dani Gamerman ·  
Marina Silva Paez

Received: 5 August 2011 / Revised: 26 June 2012 / Published online: 31 July 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** Space deformation has been proposed to model space-time varying observation processes with non-stationary spatial covariance structure under the hypothesis of temporal stationarity. In real applications, however, the temporal stationarity assumption is inappropriate and unrealistic. In this work we propose a spatial-temporal model whose temporal trend is modeled through state space models and a spatially varying anisotropy is modeled through spatial deformation, under the Bayesian approach. A distinctive feature of our approach is the consideration of model uncertainty in an unified framework. Our model has a clear advantage over the ones proposed so far in the literature when the main objective of the study is to perform spatial interpolation for fixed points in time. Approximations of the posterior distributions of the model parameters are obtained via Markov chain Monte Carlo methods. This allows for prediction of the process values in space and time as well as handling of missing values. Two applications are presented: the first one to model concentrations of sulfur dioxide in the eastern United States and the second one to model monthly minimum temperatures in the State of Rio de Janeiro.

**Keywords** Anisotropy · Bayesian inference · Concentrations of sulfur dioxide · MCMC · Minimum temperature · Spatial deformation · State space models

---

F. E. Castro Morales  
DEST, UFRN, Natal, Brazil

D. Gamerman · M. S. Paez (✉)  
DME, UFRJ, Rio de Janeiro, Brazil  
e-mail: marina@im.ufrj.br

## 1 Introduction

In geostatistics, the simplest models are the ones built under the assumption of stationarity and isotropy of the spatial covariance matrix. These assumptions, however, are not always realistic. When environmental processes are under study, for example, the observations are usually affected by specific landscape and topographic characteristics of the region of interest. In fact, in the analysis of most spatial-temporal processes there is little reason to expect spatial covariance structures to be stationary over the spatial scales of interest (Sampson et al. 2001; Sampson and Guttorp 1992). Relaxing these assumptions, Sampson and Guttorp (1992) proposed a nonparametric approach to global estimation of the spatial covariance structure of a spatial-temporal process  $Y(s_i, t)$  observed at a set of  $n$  locations in space  $\{s_1, \dots, s_n\} \in S$  and  $T$  points in time  $t = 1, \dots, T$ . In their model, they assume temporal stationarity, such that the observations can be thought of as repeated measurements of a random variable  $Y$  at the fixed set of locations  $\{s_1, \dots, s_n\}$ . To deal with non-stationarity and anisotropy in space, they work with a process of spatial deformation. The idea is to map the geographic coordinates from the original space,  $G$ -space, into a new latent space called  $D$ -space, where the hypotheses of isotropy and stationarity hold. To obtain the location of sampled monitoring stations in  $D$ -space, they made use of multidimensional scaling. After the locations are obtained in  $D$ -space, Sampson and Guttorp (1992) apply thin-plate splines for interpolation purposes.

One disadvantage of the model proposed by Sampson and Guttorp (1992) is that it does not take into account the uncertainty associated with the estimation of the spatial deformation. This problem was handled by Damian et al. (2001) and Schmidt and O'Hagan (2003), who independently proposed Bayesian versions of the model proposed by Sampson and Guttorp (1992), taking that uncertainty into consideration. The main difference between the approaches of Damian et al. (2001) and Schmidt and O'Hagan (2003) is in the specification of the prior distribution of the spatial deformation. Damian et al. (2001) proposed the use of a prior suggested by Mardia et al. (1991), which is based on the bending energy of the spline, being a natural prior for thin-plate spline mapping. Schmidt and O'Hagan (2003), on the other hand, proposed a Gaussian process as a prior for the function which maps the locations into the latent  $D$ -space.

All the methodologies presented so far assume temporal stationarity of the spatial-temporal process. This assumption can be appropriate for suitably short points in time. In practice this is not feasible and stationarity is typically achieved for residuals obtained after removing temporal trends for each monitoring station. One such approach that allows the incorporation of a temporal trend in the model was proposed by Bruno et al. (2008), extending the spatio-temporal model proposed by Bruno and Cocchi (2004) to incorporate anisotropy via the spatial deformation of Sampson and Guttorp (1992). They assume that the observation made at location  $s_i \in S$ , and time  $t$ , denoted by  $Y(s_i, t)$ , is defined by

$$Y(s_i, t) = \mu(s_i, t) + \sigma_t(s_i)Z(s_i, t) + \xi(s_i, t),$$

where  $\mu(s_i, t)$  is the spatio-temporal mean trend,  $Z(s_i, t)$  denotes a zero mean smoothed Gaussian spatiotemporal underlying process and  $\xi(s_i, t)$  is a Gaussian noise process which represents measurement errors.  $\sigma_t(s_i)$  is an extra variance component incorporated to handle possible features of seasonal variability, and it represents the part of the non-stationary temporal variability that, when removed, leaves separable space and time correlation components. The main disadvantage of this approach is the way the parameters are estimated, which consists of separate steps for preprocessing the original data. Firstly,  $\sigma_t(s_i)$  is estimated using a Box–Jenkins time series analysis (Box and Jenkins 1976). The estimated values are used to standardize the data and classical separable space-time correlation models are applied. After that, an analysis of spatial non-stationarity is performed under the Damian et al. (2001) approach for estimating heterogeneous spatial covariance.

In this work we propose a spatial-temporal model with a non-stationary spatial covariance structure, which, as in Bruno et al. (2008), allows for the incorporation of a temporal trend. However, this paper proposes a fully Bayesian approach for inference, with simultaneous estimation of all model components in a single step. We propose a model whose temporal trend is modeled through the state space models (West and Harrison 1997) and corrects anisotropy via the spatial deformation (Sampson and Guttorp 1992) under the Bayesian approach proposed by Schmidt and O'Hagan (2003). In both applications that will be presented later in this paper, the temporal trend include seasonal terms. Our model has a clear advantage over the ones proposed so far in the literature when the main objective of the study is to perform spatial interpolation for fixed points in time, as it has all the benefits shared by the models estimated under a fully Bayesian approach, like Damian et al. (2001) and Schmidt and O'Hagan (2003), but without the need of estimating the temporal trend independently from the spatial structure.

This paper is organized as follows: in Sect. 2 we introduce the proposed model; in Sect. 3 aspects of inference are presented with the specification of the prior, likelihood and posterior distributions as well as the predictive function used to perform interpolation in space; in Sect. 4 simulation studies are presented and illustrate the advantages of our approach; in Sect. 5 two applications with real data-sets are presented: one to concentrations of sulfur dioxide in the eastern United States and the other to minimum temperatures in the State of Rio de Janeiro. Finally, Sect. 6 presents the concluding remarks.

## 2 Model definition

Suppose that observations are made at  $n$  monitoring stations in a certain geographic region of interest  $G$ ,  $G \subset \mathbb{R}^2$ , over  $T$  distinct points in time. Define  $s_i = (x_{i1}, x_{i2})'$  as the geographic coordinates corresponding to the  $i$ th monitoring station,  $i = 1, \dots, n$ , and denote by  $Y(s_i, t)$  the observed value in time  $t$  at the geographic location  $s_i$ . We propose a space-time process for the observations  $Y_t = (Y(s_1, t), \dots, Y(s_n, t))'$ ,  $t = 1, \dots, T$ , specified by the sum of a mean process  $\mu_t$  and a zero mean Gaussian noise process  $\epsilon_t$ . The mean process  $\mu_t$  is specified as a state space model (West and Harrison 1997, chapter 4). It is written as a linear combination of explanatory variables, as in

a linear regression model, but the regression coefficients are allowed to vary in time. This model is composed of three parts: an observation equation, a structural equation and a system equation, which are given below:

$$\text{Observation equation: } Y_t = \mu_t + \epsilon_t, \quad (1)$$

$$\text{Structural equation: } \mu_t = F_t' \beta_t, \quad (2)$$

$$\text{System equation: } \beta_t = G_t \beta_{t-1} + \omega_t, \quad \omega_t \sim N_p(\mathbf{0}_p, \Psi), \quad (3)$$

for  $t = 1, \dots, T$ .  $N_p(\mu, \Sigma)$  denotes the  $p$ -variate normal distribution with mean vector  $\mu$  and variance-covariance matrix  $\Sigma$ ,  $F_t$  is an  $n \times p$  matrix of covariates,  $\beta_t = (\beta_{0t}, \dots, \beta_{(p-1)t})'$  is the vector of regression coefficients,  $G_t$  is a known  $p \times p$  evolution matrix,  $\mathbf{0}_p$  denotes a vector of zeros of size  $p$  and  $\Psi$  is the covariance matrix of  $\omega_t$ .

The above formulation includes much more than allowing for time-varying regression coefficients. It includes many components usually encountered in standard time series analysis such as time trends, seasonality and cycles and includes ARMA model specifications. The reader is referred to [Harvey \(1989\)](#) and [West and Harrison \(1997\)](#) for details.

The model is completed with the specification of a multivariate normal prior distribution for  $\beta_0$  and an Inverted Wishart prior distribution for the covariance matrix  $\Psi$ , denoted by

$$\beta_0 \sim N_p(m_0, C_0), \quad (4)$$

$$\Psi \sim W_{n_0}^{-1}(S_0), \quad (5)$$

where  $S_0$  is the mean and  $n_0$  is the degrees of freedom of the Inverted Wishart prior.

A spatial structure is incorporated in the model via the noise process  $\epsilon_t = (\epsilon(s_1, t), \dots, \epsilon(s_n, t))'$ , through the specification of a spatially structured variance-covariance matrix. Assume that the covariance between elements  $\epsilon(s_i, t)$  and  $\epsilon(s_j, t)$ ,  $s_i \neq s_j$ ,  $s_i, s_j \in G$ , do not depend on the time  $t$ , and  $\omega_t$  and  $\epsilon_t$  are independent, for all  $t$ . We define the covariance function for  $\epsilon_t$  by

$$\text{Cov}(\epsilon(s_i, t), \epsilon(s_j, t)) = \sqrt{v_i v_j} \rho_b(|d(s_i) - d(s_j)|), \quad (6)$$

where  $v_i$  is the variance of  $\epsilon(s_i, t)$ , which is assumed constant for every period of time  $t$ , for  $i = 1, \dots, n$ ;  $\rho_b$  is a valid isotropic correlation function, which depends on parameter  $b$ .

The function  $d(\cdot)$  maps the original geographic coordinates in  $G$ -space to coordinates in the new latent space  $D$ .

Note that to model observations under the hypothesis of isotropy, the function  $d(\cdot)$  must be specified as the identity function. In this work we do not assume isotropy and apply the space deformation idea of Sampson and Guttorp with a Gaussian Process prior distribution for  $d(\cdot)$ , as proposed by [Schmidt and O'Hagan \(2003\)](#):

$$d(\cdot) \sim GP(g(\cdot), \sigma_d^2, \rho_\phi), \quad (7)$$

where

- $g(\cdot)$  is the prior mean function, which can take any form that reflects the prior beliefs about the mapping from spaces  $G$  to  $D$ . Note that this is a function of the spatial coordinates of the locations in space  $G$ . In this paper we assume that  $g(\mathbf{x}) = \mathbf{x}$ , which is a reasonable assumption when there are no prior beliefs about how  $D$  should differ from  $G$ .
- $\sigma_d^2$  is a matrix of variances which, for  $G \in \mathbb{R}^2$ , have dimension  $2 \times 2$ . Here we define  $\sigma_d^2$  as the diagonal matrix  $\sigma_d^2 = \text{diag}(\sigma_{d_{11}}^2, \sigma_{d_{22}}^2)$ , as suggested by Schmidt and O'Hagan (2003) to avoid unidentifiability of the model. Larger elements of  $\sigma_d^2$  allow greater distortions from  $G$  to  $D$  space.
- $\rho_\phi$  is any valid correlation function which measures the prior correlation between locations in  $D$ -space, and  $\phi$  is a known set of parameter, such that

$$\text{Cor}(d(s_i), d(s_j)) = \begin{cases} \rho_\phi(|s_i - s_j|), & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$$

In particular, let  $\mathbf{d}^G = (\mathbf{d}_1, \dots, \mathbf{d}_n)$ , where  $\mathbf{d}_i = d(s_i)$ , be the  $2 \times n$  matrix of the coordinates of the gauged locations in  $D$ -space; define  $\mathbf{g} = (g(s_1), \dots, g(s_n))$ ; and let  $\mathbf{R}_d^G$  be the  $n \times n$  matrix of correlations such that  $\mathbf{R}_d^G[i, j] = \text{Cor}(d(s_i), d(s_j))$ . Then the prior for  $\mathbf{d}^G$  is a Matrix Normal distribution with mean  $\mathbf{g}$ , row covariance matrix  $\sigma_d^2$  and column covariance matrix  $\mathbf{R}_d^G$ , denoted by:

$$\mathbf{d}^G \sim N_{(2 \times n)}(\mathbf{g}, \sigma_d^2, \mathbf{R}_d^G). \quad (8)$$

It is also important to notice that the function  $d(\cdot)$  is unidentifiable, as all the transformations of the coordinates in  $D$ -space which keep the same distances between locations are observationally equivalent. Besides that, any transformation that multiplies all distances in  $D$ -space by a constant is equivalent. To guarantee the identifiability of function  $d(\cdot)$  we fix the location of two sites in  $D$ -space. This restriction plays the role of an anchor, guaranteeing that  $d(\cdot)$  will not rotate or translate. In applications, it is recommended to set that the two known locations in  $D$ -space equal their corresponding geographic locations in  $G$ -space.

### 3 Inference

#### 3.1 Prior distributions

Under the Bayesian point of view, the model specified in the previous section must be completed with the prior specification of the parameters  $\mathbf{v} = (v_1, \dots, v_n)$  and  $b$  (in 7) and  $\sigma_d^2$  (in 9). Here we consider the same prior distributions proposed by Schmidt and O'Hagan (2003) for the parameters  $\mathbf{v}$  and  $\sigma_d^2$ , and the one proposed by Schmidt and Gelfand (2003) for the parameter  $b$ . These priors distributions are given as follows:

- Assuming that  $\mathbf{v} = (v_1, \dots, v_n)$  are exchangeable, an Inverted Gamma prior distribution for  $v_i, i = 1, \dots, n$ , is specified as:

$$v_i \mid \tau^2, f \sim \text{IG}(f, \tau^2(f - 2)),$$

where  $f$  is the number of degrees of freedom, considered known, and  $\tau^2$  is the mean of the distribution, which has a non-informative prior distribution given by:

$$\pi(\tau^2) \propto (\tau^2)^{-1}$$

- We assume that the elements of the main diagonal of  $\sigma_d^2$  are independent with an Inverted Gamma prior distribution for each element, given by

$$\sigma_{d_{ii}}^2 \sim \text{IG}(\eta_i, \psi_i), \quad \text{for } i = 1, 2.$$

- The prior distribution for  $b$  depends on the choice of correlation function for the observational error structure. When  $b$  is a scalar representing the range, as in the exponential correlation function  $\rho_b(x) = \exp\{-bx\}$ , an informative Gamma distribution

$$b \sim G(b^* \eta, \eta),$$

where  $b^* = -2 \log(0.05) / \max(|s_i - s_j|)$  was specified.

That specification assumes that a priori the correlation of observations in two remote locations in the region of study is close to 0.

Note that the prior distributions for the parameters  $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \dots, \boldsymbol{\beta}_T)'$ ,  $\mathbf{d}^G$  and  $\boldsymbol{\Psi}$  were previously specified in (4), (5), (6) and (7). The prior distribution for the set of all of the unknown parameters of the model  $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \mathbf{d}^G, \mathbf{v}, b, \tau^2, \sigma_d^2, \boldsymbol{\Psi}\}$ , under the hypothesis of independence between them, is given by

$$\pi(\boldsymbol{\theta}) = \left\{ \prod_{t=1}^T \pi(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1}, \boldsymbol{\Psi}) \right\} \pi(\boldsymbol{\beta}_0) \pi(\boldsymbol{\Psi}) \pi(\mathbf{d}^G) \left\{ \prod_{i=1}^n \pi(v_i \mid \tau^2) \right\} \pi(\tau^2) \pi(b) \pi(\sigma_d^2). \quad (9)$$

### 3.2 Posterior distribution

Suppose that observations  $\mathbf{y}_t$  from the model in (1)–(3) were made for  $t = 1, \dots, T$ . Define a matrix of observations  $\mathbf{y}$  as  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$ . The likelihood function of  $\boldsymbol{\theta}$  given  $\mathbf{y}$  is

$$L(\boldsymbol{\theta} \mid \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \mathbf{F}_t' \boldsymbol{\beta}_t)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{F}_t' \boldsymbol{\beta}_t) \right\}. \quad (10)$$

The combination of the likelihood function in (10) with the prior distribution for  $\theta$ , given in (9), gives via Bayes theorem the posterior distribution of  $\theta$

$$\begin{aligned} \pi(\theta \mid y) \propto L(\theta \mid y) & \left\{ \prod_{t=1}^T \pi(\beta_t \mid \beta_{t-1}, \Psi) \right\} \pi(\beta_0) \pi(\Psi) \pi(d^G) \\ & \times \left\{ \prod_{i=1}^n \pi(v_i \mid \tau^2) \right\} \pi(\tau^2) \pi(b) \pi(\sigma_d^2). \end{aligned} \quad (11)$$

The posterior distribution in (11) does not have a closed form and we use MCMC methods to obtain approximations for this distribution. The algorithm used to obtain samples from the posterior distribution in (11) is a hybrid algorithm, with the use of *Gibbs sampling* for the parameters whose full conditional density is known, and *Metropolis-Hastings* (Metropolis et al. 1953; Hastings 1970) steps to sample from the other parameters. The computational aspects of this work can be found in details in Castro (2010). More details about MCMC methods can be seen in Gamerman and Lopes (2006) and Robert and Casella (2004).

### 3.3 Interpolation

One of the main interests in geostatistics is to perform interpolation, which means to obtain predictions of the process under study for non monitored (or ungauged) locations in space. In the proposed model, to interpolate the response variable  $Y_t$  for a fixed period of time  $t$ , one must first interpolate the deformations of the geographic locations of interest. In this section we show how to add additional steps in the MCMC algorithm to interpolate the deformations and then interpolate  $Y_t$ .

- Interpolation of the deformations of the geographic locations

Let  $s^G = (s_1, \dots, s_n)$  be the matrix of coordinates of the gauged geographic locations and  $s^{UG} = (s_{n+1}, \dots, s_{n+m})$  be a matrix of coordinates of  $m$  ungauged geographic locations.  $d^G = (d_1, \dots, d_n)$ , as previously specified, is the matrix of coordinates of the gauged locations in  $D$ -space. Define  $d^{UG} = (d_{n+1}, \dots, d_{n+m})$  as the matrix of coordinates of the ungauged locations in  $D$ -space, and  $d = (d^G, d^{UG})$ . Denote by  $v^{UG} = (v_{n+1}, \dots, v_{n+m})$  the vector of variances in the process at ungauged locations. Noting that  $d$  follows a Gaussian process as specified in (7), we have that the prior distribution for  $d$  is a Matrix Normal distribution analogous to (8):

$$(d^G, d^{UG}) \mid R_d, \sigma_d^2 \sim N_{(2 \times (n+m))} \left( (s^G, s^{UG}), \sigma_d^2, \begin{pmatrix} R_d^G & R_d^{*'} \\ R_d^* & R_d^{UG} \end{pmatrix} \right), \quad (12)$$

where  $R_d^G$  is the matrix of correlations between the gauged locations;  $R_d^{UG}$  is the matrix of correlations between the ungauged locations;  $R_d^*$  is the matrix of correlations between gauged and ungauged locations; and  $R_d = \begin{pmatrix} R_d^G & R_d^{*'} \\ R_d^* & R_d^{UG} \end{pmatrix}$ .

That way, using properties of the Multivariate Normal distribution we have that,

$$\text{vec}(\mathbf{d}^{UG}) \mid \mathbf{d}^G, \mathbf{R}_d, \sigma_d^2 \sim N_{2m}(\mathbf{A}, \mathbf{B}), \quad (13)$$

where  $\mathbf{A} = \text{vec}(\mathbf{s}^{UG}) + (\mathbf{I}_2 \otimes \mathbf{R}_d^*(\mathbf{R}_d^G)^{-1})(\text{vec}(\mathbf{d}^G) - \text{vec}(\mathbf{s}^G))$ ,  $\mathbf{B} = \sigma_d^2 \otimes (\mathbf{I}_m - \mathbf{R}_d^*(\mathbf{R}_d^G)^{-1}\mathbf{R}_d^*)$ ,  $\text{vec}(\mathbf{X})$  denotes the vectorization form of the matrix  $\mathbf{X}$  and  $\mathbf{I}_n$  denotes the identity matrix of order  $n$ .

Samples of the posterior distribution of  $\mathbf{d}^{UG}$  and  $\mathbf{v}^{UG}$  can be easily obtained by adding steps in the MCMC algorithm. For the  $k$ th iteration of the algorithm, given the sampled values for  $\theta$ , say  $\theta^{(k)}$ , the parameters  $\mathbf{v}^{UG(k)}$  and  $\mathbf{d}^{UG(k)}$  are sampled from

$$v_j \mid \theta^{(k)} \sim IG(\tau^{2(k)}(f-2), f), \quad (14)$$

for  $j = n+1, \dots, n+m$  and

$$\text{vec}(\mathbf{d}^{UG(k)}) \mid \theta^{(k)} \sim N_{2m}(\mathbf{A}^{(k)}, \mathbf{B}^{(k)}), \quad (15)$$

where  $\mathbf{A} = \text{vec}(\mathbf{s}^{UG}) + (\mathbf{I}_2 \otimes \mathbf{R}_d^*(\mathbf{R}_d^G)^{-1})(\text{vec}(\mathbf{d}^{G(k)}) - \text{vec}(\mathbf{s}^G))$  and  $\mathbf{B} = \sigma_d^{2(k)} \otimes (\mathbf{I}_m - \mathbf{R}_d^*(\mathbf{R}_d^G)^{-1}\mathbf{R}_d^*)$ .

#### • Interpolation of $\mathbf{Y}$

Denote by  $\mathbf{Y}_t^{UG} = (Y(s_{(n+1)}, t), \dots, Y(s_{(n+m)}, t))$  the vector of response variables in time  $t$ ,  $t = 1, \dots, T$ , for the ungauged geographic locations, and  $\mathbf{Y}_t^A = (\mathbf{Y}_t, \mathbf{Y}_t^{UG})$  the vector of response variable at gauged and ungauged locations for a fixed period of time  $t$ . Defining  $\theta^A = (\theta, \mathbf{v}^{UG}, \mathbf{d}^{UG})$  we have that

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_t^{UG} \end{pmatrix} \mid \theta^A \sim N \left[ \begin{pmatrix} \mathbf{F}_t' \boldsymbol{\beta}_t \\ \mathbf{F}_t'^{UG} \boldsymbol{\beta}_t \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}^* \\ \boldsymbol{\Sigma}^* & \boldsymbol{\Sigma}^{UG} \end{pmatrix} \right],$$

where  $\mathbf{F}_t'$  is the dynamic regression matrix for the gauged locations,  $\mathbf{F}_t'^{UG}$  is the dynamic regression matrix for the ungauged locations,  $\boldsymbol{\Sigma}$  is the covariance matrix for the gauged locations,  $\boldsymbol{\Sigma}^{UG}$  is the covariance matrix for the ungauged locations and  $\boldsymbol{\Sigma}^*$  is the covariance matrix between gauged and ungauged locations. Using basic properties of the Multivariate Normal distribution, we have that:

$$\mathbf{Y}_t^{UG} \mid \theta^A, \mathbf{y}_t \sim N(\mathbf{F}_t'^{UG} \boldsymbol{\beta}_t + \boldsymbol{\Sigma}^{*'}(\boldsymbol{\Sigma})^{-1}(\mathbf{y}_t - \mathbf{F}_t' \boldsymbol{\beta}_t), \boldsymbol{\Sigma}^{UG} - \boldsymbol{\Sigma}^{*'}(\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^*).$$

The density of the predictive distribution of  $\mathbf{Y}_t^{UG}$  given the observations  $\mathbf{y}$  is given by

$$\pi(\mathbf{y}_t^{UG} \mid \mathbf{y}) = \int_{\theta^A} \pi(\mathbf{y}_t^{UG} \mid \mathbf{y}_t, \theta^A) \pi(\theta^A \mid \mathbf{y}) d\theta^A. \quad (16)$$



Note that the integral in (16) is not analytically tractable. We use Monte Carlo methods to obtain an approximation of (16) as follows:

$$\begin{aligned}\pi_j(\mathbf{y}_t^{UG} | \mathbf{y}) &= \int_{\boldsymbol{\theta}^A} \pi(\mathbf{y}_t^{UG} | \mathbf{y}_t, \boldsymbol{\theta}^A) \pi(\boldsymbol{\theta}^A | \mathbf{y}) d\boldsymbol{\theta}^A, \\ &= \int_{\boldsymbol{\theta}^A} \pi(\mathbf{y}_t^{UG} | \mathbf{y}_t, \boldsymbol{\theta}^A) \pi(\mathbf{v}^{UG}, \mathbf{d}^{UG} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}^A, \\ &\approx \frac{1}{K} \sum_{k=1}^K \pi(\mathbf{y}_t^{UG} | \boldsymbol{\theta}^{A(k)}, \mathbf{y}_t),\end{aligned}$$

where  $\boldsymbol{\theta}^A$  is sampled from its prior distribution  $\pi(\boldsymbol{\theta}^A | \mathbf{y})$  in the following way: firstly,  $\boldsymbol{\theta}^{(k)}$  is sampled from  $\pi(\boldsymbol{\theta} | \mathbf{y})$ , then  $\mathbf{v}^{UG(k)}$  and  $\mathbf{d}^{UG(k)}$  are sampled from  $\pi(\mathbf{v}^{UG(k)}, \mathbf{d}^{UG(k)} | \boldsymbol{\theta}^{(k)})$ , for  $k = 1, \dots, K$ .

## 4 Simulation studies

In this section we investigate properties of the proposed methodology through the use of an artificial data-set simulated over  $n = 15$  locations in space and  $T = 100$  points in time. The data was simulated from a dynamic model which takes into account a seasonal effect and the effect of an artificial 0–1 covariate  $\mathbf{x}$ , fixed in time. The model is given as follows:

$$\begin{aligned}\mathbf{Y}_t &= \mathbf{F}_t' \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N_n(\mathbf{0}_n, \boldsymbol{\Sigma}), \\ \boldsymbol{\beta}_t &= \mathbf{G}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N_p(\mathbf{0}_p, \boldsymbol{\Psi}),\end{aligned}$$

where,  $\boldsymbol{\beta}_t = (\beta_{t0}, \beta_{t1}, \beta_{t2}, \beta_{t3})'$ ,  $\mathbf{F}_t' = (\mathbf{1}_n, \mathbf{x}, \mathbf{1}_n, \mathbf{0}_n)$ ,  $\mathbf{x} = (\mathbf{0}_{10}, \mathbf{1}_{17})'$ ,  $\mathbf{G}_t = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{G}_2 \end{pmatrix}$ ,  $\mathbf{G}_2 = \begin{pmatrix} \cos(\pi)/25 & \sin(\pi)/25 \\ -\sin(\pi)/25 & \cos(\pi)/25 \end{pmatrix}$ ,  $\boldsymbol{\Psi} = \text{diag}(0.01, 0.02, 0.06, 0.06)$  and  $\boldsymbol{\beta}_0 = (0.01, 3, 5, 1)$ . The 15 locations  $\mathbf{d}_i$ 's were simulated from (8) with a Gaussian correlation function  $\rho_\phi$  with parameter  $\phi = 3.28$  ( $\phi = -2 \log(0.05) / \max_{i,j=1,\dots,n} (|s_i - s_j|)^2$ ) and  $\sigma_{d_{ii}}^2 = 0.015$ ,  $i = 1, 2$ . An exponential correlation function with  $b = 6$  was assumed for the observation errors.

### 4.1 Sensitivity study for prior specifications

In this section a sensitivity analysis of the model was performed to assess the impact of different prior specifications for the parameters  $b$  and  $\sigma_d^2$ . Three different prior distributions for  $b$  and for  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$  were considered. For the parameter  $b$ , Gamma prior distributions were specified with the same mean (equal to the true value), and different variances, aiming to compare different degrees of information in the prior distribution. The three specifications are given below:

- b1.  $b \sim G(52 \times 6, 52),$
- b2.  $b \sim G(10 \times 6, 10),$
- b3.  $b \sim G(1 \times 6, 1).$

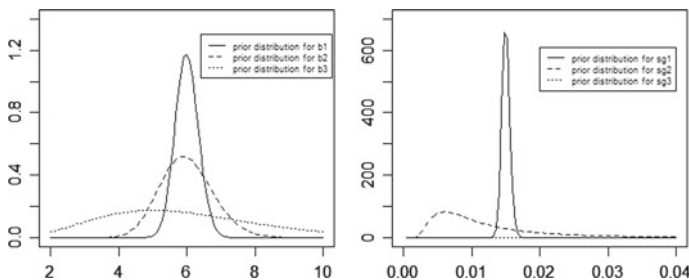
For the parameters  $\sigma_{dii}^2, i = 1, 2,$  three Inverted Gamma prior specifications were made:

- sg1.  $\sigma_{dii}^2 \sim IG(66.7 \times 10, 10), \quad i = 1, 2,$
- sg2.  $\sigma_{dii}^2 \sim IG(66.7 \times 0.02 + 1, 0.02), \quad i = 1, 2,$
- sg3.  $\sigma_{dii}^2 \sim IG(0.01, 0.01), \quad i = 1, 2.$

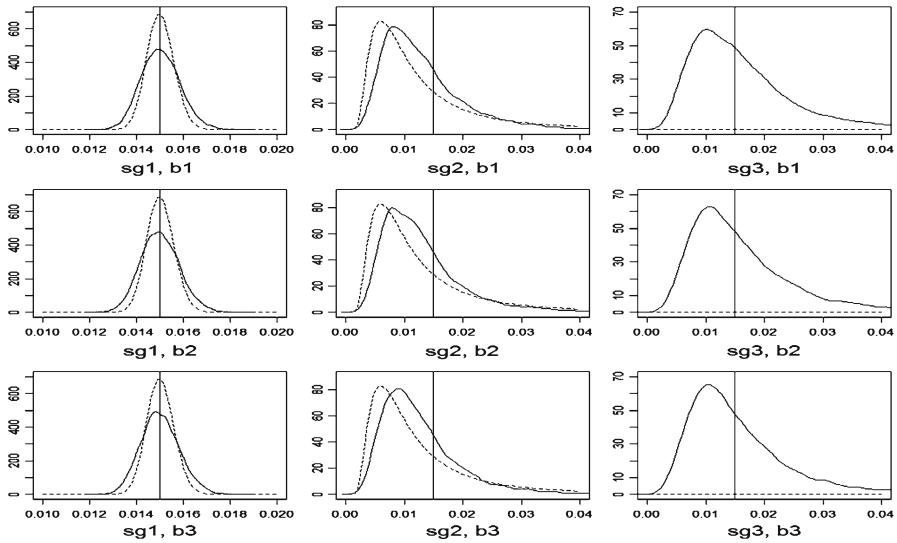
In the first two specifications for  $\sigma_{dii}^2, i = 1, 2,$  the mean is equal to the true variance (approximately equal to 0.015) with different degrees of precision (the prior variances are, respectively  $3.39 \times 10^7$  and 0.00067). The third specification is a reasonably vague prior distribution (Berger et al. 2001; Gelman 2006). That way, we considered nine different models obtained by the combination of the three prior specifications suggested for the parameter  $b$  and the three prior specifications suggested for the parameter  $\sigma_d^2$ .

Figure 1 shows the comparison between the different prior specifications for the parameters  $b$  and  $\sigma_{dii}^2, i = 1, 2.$  The different amount of information provided by each specification is clear from the figure, with the first specification in both cases, being highly informative, the third being vaguely informative, and the second, in between. A graphical representation of the posterior distribution for  $b$  and  $\sigma_{dii}^2, i = 1, 2,$  under each one of the nine combinations of prior distribution, can be seen, respectively in Figs. 2 and 3. It can be seen that the posterior densities for  $\sigma_{d11}^2$  and  $b$  are rather insensitive to changes in the prior specification. The only exception is when using the more informative prior specification for  $\sigma_{d11}^2$ , which does affect the form of the posterior distribution of this parameter.

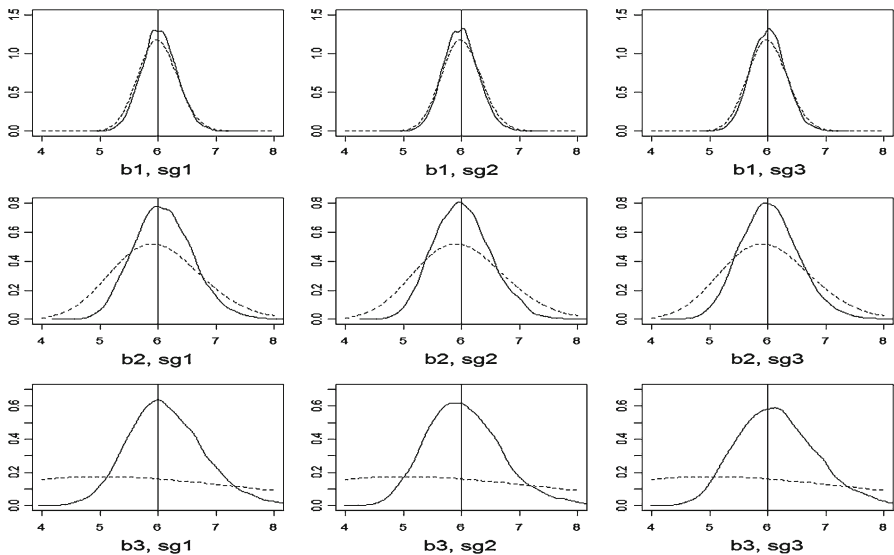
Figure 4 compares the true deformation and the estimated deformation (posterior mean) under all the combinations of prior distributions for the parameters  $\sigma_{d11}^2$  and  $b.$  It can be seen that the estimated deformation remains almost the same regardless of the choice of priors, showing that the estimation of the deformation is not sensitive to the changes in the priors for  $\sigma_{d11}^2$  and  $b.$



**Fig. 1** Prior distributions for the parameter  $b$  (on the left hand side) and the parameters  $\sigma_{dii}^2, i = 1, 2$  (on the right hand side)

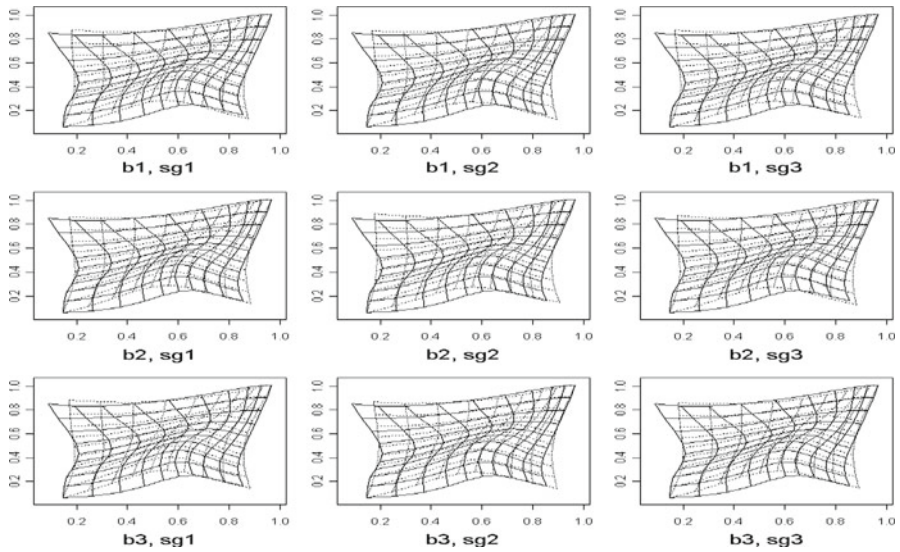


**Fig. 2** Posterior distribution (solid line) and prior distribution (dashed line) for the parameter  $\sigma^2_{d11}$  under all the combinations of prior distributions for  $\sigma^2_{d11}$  (varying by column) and  $b$  (varying by row)



**Fig. 3** Posterior distribution (solid line) and prior distribution (dashed line) for the parameter  $b$  under all the combinations of prior distributions for  $\sigma^2_{d11}$  (varying by column) and  $b$  (varying by row)

We use the *DIC* statistics (Spiegelhalter et al. 2002) to compare the nine fitted models. The *DIC* is a Bayesian version of the *AIC* (Akaike Information Criterion), and like other widely used model comparison methods (such as *AIC* and *BIC*—Bayesian Information Criterion), the *DIC* can be thought of a goodness of fit measurement



**Fig. 4** True deformation (*solid line*) and estimated deformation (*dashed line*) under all the combinations of prior distributions for  $\sigma_{d11}^2$  (varying by *column*) and  $b$  (varying by *row*)

**Table 1** DIC statistics for models with different prior specifications for the parameters  $b$  and  $\sigma_d^2$

	sg1	sg2	sg3
b1	542.55	544.41	545.64
b2	543.92	547.60	545.76
b3	545.04	546.23	547.11

which is penalized by the effective number of parameters in the model. Models with smaller *DIC* should be preferred. Table 1 presents the *DIC* statistics for the fitted models. The gain obtained by using more informative priors for the parameters  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$  and  $b$  is small. Even when using vaguely informative priors the model can estimate the true values of these parameters well, as seen by Figs. 2 and 3. Therefore, the main conclusion taken from this sensitivity study is that when the true spatial deformation is smooth, the posterior distribution of  $\theta$  is similar when using vague and informative priors for  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ , and  $b$  centered around their true value.

#### 4.2 Study of the uncertainty of the spatial deformation

The inference procedure performed for the spatial deformation is based on the posterior distribution of  $\theta$ , which does not have a closed form. In this section, we propose an approximate credibility region for  $d_i$  based on asymptotic theory (Gamerman and Migon 1999; Casella and Berger 2001), and we perform a study about the uncertainty of the estimation of this deformation for the simulated data-set.

The proposed approximation uses the fact that  $[d_i | y]$  is approximately distributed as  $N(\hat{d}_i, J^{-1}(\hat{d}_i))$ , for large  $T$  where

$$J^{-1}(\hat{d}_i) = -\frac{\partial^2 \pi(\boldsymbol{\theta} \mid \mathbf{y})}{\partial \mathbf{d}_i \mathbf{d}_i'}.$$

That way, the  $100(1 - \alpha)\%$  asymptotic credibility region  $C \subset \mathbb{R}^2$  is such that

$$P \left[ (\mathbf{d}_i - \hat{\mathbf{d}}_i) J^{-1}(\hat{\mathbf{d}}_i) (\mathbf{d}_i - \hat{\mathbf{d}}_i) \in C \right] \geq 1 - \alpha. \quad (17)$$

It is quite complicated to obtain an analytical expression for  $\hat{\mathbf{d}}_i$  or  $J^{-1}(\hat{\mathbf{d}}_i)$ . Fortunately, these quantities can be well approximated by

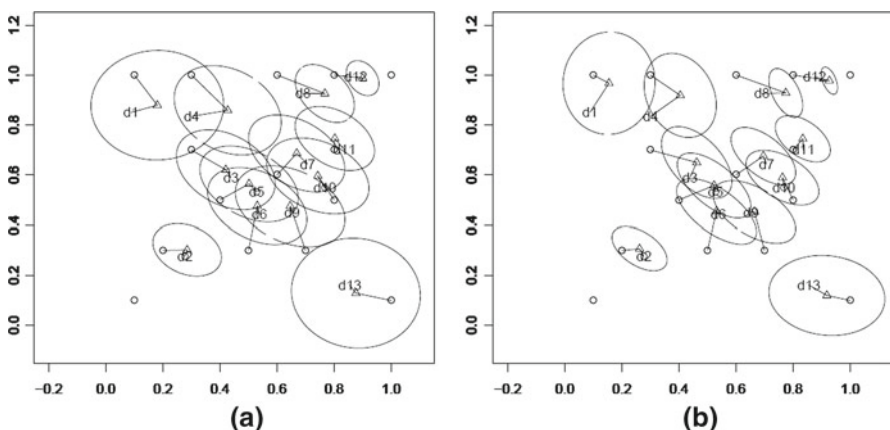
$$\hat{\mathbf{d}}_i \approx \frac{1}{K} \sum_{j=1}^K \mathbf{d}_i^{(j)}, \text{ and } J^{-1}(\hat{\mathbf{d}}_i) \approx \frac{1}{K-1} \sum_{j=1}^K (\mathbf{d}_i^{(j)} - \hat{\mathbf{d}}_i)(\mathbf{d}_i^{(j)} - \hat{\mathbf{d}}_i)',$$

where  $\mathbf{d}_i^{(j)}$  are posterior samples from  $\mathbf{d}_i$  obtained via MCMC.

Figure 5 shows the estimated 95 % credibility regions for each  $\mathbf{d}_i$ ,  $i = 1, \dots, 15$ , with  $T = 100$  and  $T = 500$ . It can be seen that the credibility regions are larger when working with  $T = 100$ , showing that when we increase the sample size the precision in the estimation of the parameters improves.

### 4.3 Comparison against existing models

In this section we aim to compare, through simulation, the performance of our model (call it model A) against other models proposed in the literature, when the observed process presents anisotropy. The first comparison is made with a simplified version of the proposed model, considering the same dynamic structure but assuming the deformation  $d(\cdot)$  to be fixed as the identity (call it model B). Model B would be appropriate



**Fig. 5** Approximate 95 % credibility region for the parameters  $\mathbf{d}_i$ ,  $i = 1, \dots, 15$ . The posterior mean of the parameters  $\mathbf{d}_i$ ,  $i = 1, \dots, 15$ , represented by the symbol  $\Delta$ , was estimated with **a**  $T = 100$ ; **b**  $T = 500$ . The symbol  $\circ$  represents the original geographic coordinates

to model isotropic processes, and it is interesting to compare how it performs against the proposed model under anisotropy.

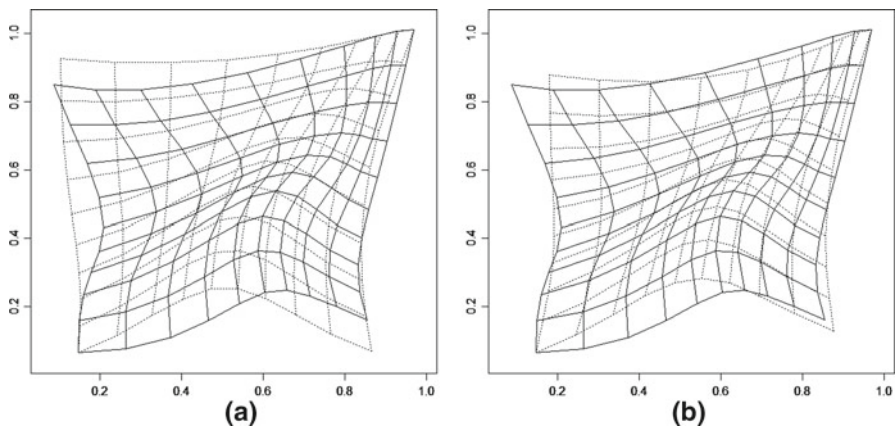
We work with the same simulated data-set used in the previous exercises but also use other simulations, varying the number of observations in time ( $T = 100$  or  $T = 500$ ) and space ( $n = 15$  or  $n = 30$ ). Table 2 shows the *DIC* statistics obtained under each simulated data-set under models *A* and *B*. The comparison shows a clear advantage of model *A*, as expected, specially when using  $n = 30$ . This result shows that with the increase of information in space, and consequently the better estimation of the spatial deformation under model *A*, the advantage of using a model which takes anisotropy into account, when the observed process is actually anisotropic, increases.

The second comparison is made between the proposed model *A* and the model proposed by Schmidt and O'Hagan (2003). As Schmidt and O'Hagan's model assume independent observation in time, for a fair comparison the model was applied to the residuals of a dynamic model with the same temporal structure as model *A*, but with no spatial structure at all. We named it model *C*.

Both methodologies take anisotropy into account. Figure 6 compares the true and estimated deformation surfaces under models *A* and *C*, and visually the estimated surface under Model *A* appears to be closer to the true deformation surface. To confirm this result, we compare the two models through the smaller Riemann distance (Kendall

**Table 2** Comparison of Models *A* and *B*, through the *DIC* statistics, under simulated data-sets with different specifications for  $n$  and  $T$

	Model <i>A</i>	Model <i>B</i>
$T = 100$		
$n = 15$	542.55	590.32
$n = 30$	110.31	228.33
$T = 500$		
$n = 15$	2,832.48	3,021.33
$n = 30$	618.72	1,218.72



**Fig. 6** True deformation (*solid line*) and estimated deformation (*dashed line*) under Models *C* (a) and *A* (b)

**Table 3** Comparison of the estimation of the spatial deformation through Models A and C using the Riemann distance

	Model A	Model C
$T = 100$		
$n = 15$	0.1148	0.1300
$n = 30$	0.0725	0.1321
$T = 500$		
$n = 15$	0.1112	0.1484
$n = 30$	0.0744	0.0911

1984) between the true spatial deformation and the ones estimated by each one of the two models. Table 3 shows the Riemann distance calculated under each simulated data-set and model, and it shows an advantage of model A, which leads to smaller distances under all scenarios. This result indicates that our model (model A) would have provided better interpolation in space than the model proposed by Schmidt and O'Hagan (2003) (model C) under the simulated examples presented here. It is important to notice, however, as already mentioned before, that the main advantage of the model proposed by us when compared to the model proposed by Schmidt and O'Hagan (2003) is methodological, as here we jointly estimate the mean of the process and the spatial covariance structure.

## 5 Applications

### 5.1 Concentrations of sulfur dioxide in the eastern United States

In this section the proposed model is applied to concentration levels of sulfur dioxide ( $\text{SO}_2$ ) measured in the eastern United States. The region of study comprises the states of Indiana, Ohio, Tennessee, Kentucky, North Carolina, Virginia, West Virginia, Pennsylvania, Maryland, New Jersey and New York. Levels of  $\text{SO}_2$  (in  $\mu\text{g}/\text{m}^3$ ) are weekly measured at 23 monitoring stations (see Fig. 9a), from the year 2000 to 2003, for a total of  $T = 208$  observations per monitoring station. To validate the performance of the proposed model, the model was fit to  $n = 22$  monitoring stations ( $s_1, \dots, s_{22}$ ), and predictions were made for station  $s_{23}$  over time, and compared to the observed values. The data-set was made available by the *Clean Air Status and Trends Network* (CASTNet), which belongs to the Environmental Protection Agency (EPA) United States, through the URL <http://www.epa.gov/castnet/>.

In this example we propose a model for the logarithm of the concentrations of  $\text{SO}_2$ , using as explanatory variables latitude (LAT), longitude (LONG), and a seasonal factor with a period of 52 weeks, corresponding to one year. The proposed model allows the regression coefficients to change smoothly over time, so that the effects of latitude and longitude are not forced to be fixed. This model is given by

$$\begin{aligned}
 Y_t &= F_t' \beta_t + \epsilon_t, \quad \epsilon_t \sim N_n(\mathbf{0}_n, \Sigma), \\
 \beta_t &= G_t \beta_{t-1} + \omega_t, \quad \omega_t \sim N_p(\mathbf{0}_p, \Psi), \\
 \beta_0 &\sim N_p(\mathbf{0}_p, 5I_p), \\
 \Psi &\sim W_{10}^{-1}(0.01I_p),
 \end{aligned}$$

where  $t = 1, \dots, 208$ ,  $F'_t = (\mathbf{1}_n, \mathbf{LONG}, \mathbf{LAT}, \mathbf{1}_n, \mathbf{0}_n)$  is a matrix of dimension  $n \times p$ , where  $\mathbf{1}_n$  denotes a vector of 1's of size  $n$ ,  $\beta_t = (\beta_0, \dots, \beta_4)'_t$  and  $\mathbf{LONG}$  and  $\mathbf{LAT}$  are the vectors of latitude and longitude corresponding to the  $n$  monitoring locations.  $G_t = \begin{pmatrix} I_3 & \mathbf{0}_{2,2} \\ \mathbf{0}_{3,3} & G_2 \end{pmatrix}$ ,  $G_2 = \begin{pmatrix} \cos \frac{\pi}{26} & \sin \frac{\pi}{26} \\ -\sin \frac{\pi}{26} & \cos \frac{\pi}{26} \end{pmatrix}$ , and  $\mathbf{0}_{k,l}$  denotes a  $k \times l$  matrix of 0's. The last two columns of  $F'_t$  are required for the representation of the time-varying seasonal component  $\beta_{3,t}$ .

Here, the elements of the matrix  $\Sigma$  are defined by

$$\text{Cov}(\epsilon(s_i, t), \epsilon(s_j, t)) = \sqrt{v_i v_j} \rho_b(|d(s_i) - d(s_j)|),$$

$$d(\cdot) \sim PG(g(\cdot), \sigma_d^2, \rho_\phi),$$

where  $g$  is the identity function and  $\rho_\phi$  is a Gaussian correlation function, with  $\phi = 4.39$  ( $\phi = -2 \log(0.05) / \max_{i,j=1,\dots,22}(|s_i - s_j|)^2$ ). The model is completed with the specification of the prior distributions for the parameters  $\mathbf{v}$ ,  $\tau^2$ ,  $b$  and  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ . We specified vaguely informative priors for the parameters  $\mathbf{v}$ ,  $\tau^2$  and informative priors for  $b$  and  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ .

The prior distribution for  $b$  was chosen to be of the form  $G(\eta b^*, \eta)$ , where  $b^* = -2 \log(0.05) / \max_{i,j=1,\dots,22}(|s_i - s_j|)$ , giving a prior expectation of  $b^* = 6$ .  $\eta$  was chosen to be equal to 52 such that the prior variance is  $2\eta b^* / \eta^2 = 0.12$ .

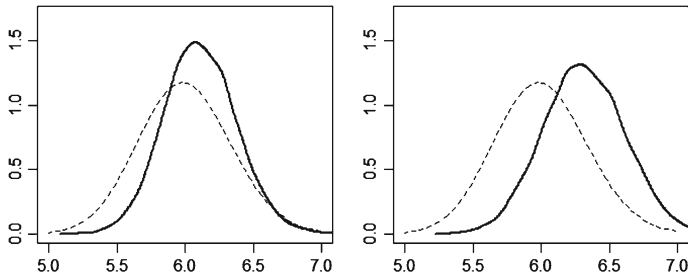
Two distinct Inverted Gamma prior distributions were proposed, for comparison, to the parameters  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ . The hyperparameters of the prior distributions were chosen according to the scale used to measure distances between locations. In this example, the scale of measurement is kilometer per 1,000 (km/1,000), so that 0.001 is equivalent to 1 km. In the first prior specification for the parameters  $\sigma_{d_{11}}^2 = \sigma_{d_{22}}^2$ , the mean was set to be equal to 0.001, with square root equal to 0.032. This means that a priori the original location can move in either direction between 0 km and  $3 \times 0.032 \times 1000 = 96$  km, with 0.997 probability. The second prior specification for  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ , was chosen to be less informative, with the prior mean being equal to 0.01. The two specifications are presented below:

1.  $\sigma_{d_{ii}}^2 \sim IG(1000 \times 100, 100)$ ,  $i = 1, 2$ ,
2.  $\sigma_{d_{ii}}^2 \sim IG(100 \times 0.01 + 1, 0.01)$ ,  $i = 1, 2$ .

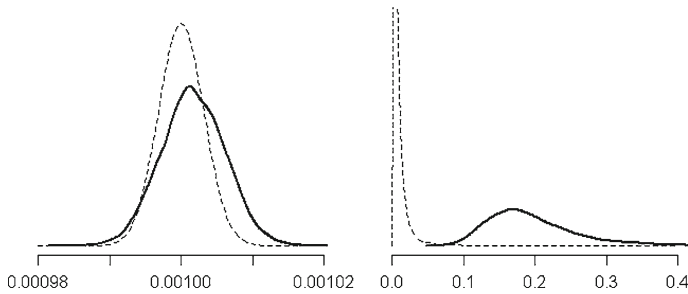
Figure 7 shows that the posterior distribution of  $b$  is not affected by changing the prior specification of  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ . The posterior distribution of  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ , however, changes significantly when changing its prior, as can be seen in Fig. 8. The posterior distribution for  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ , is similar to its prior distribution when the first prior specification is used. This effect is not observed when the second prior specification is used.

Figure 9 shows the monitoring stations in the eastern United States, with latitude and longitude converted to UTM and divided by 1,000. It also shows the estimated posterior mean of  $\mathbf{d}$  and the mapping of a regular grid of 121 points in the geographic region of study ( $G$ -space) to  $D$ -space under the first prior specifications for  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ . The MCMC chains do not seem to achieve convergence when the second prior specification is used, emphasizing the need for informative distributions for the parameters  $b$  and  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ , when it is required to account for anisotropy.

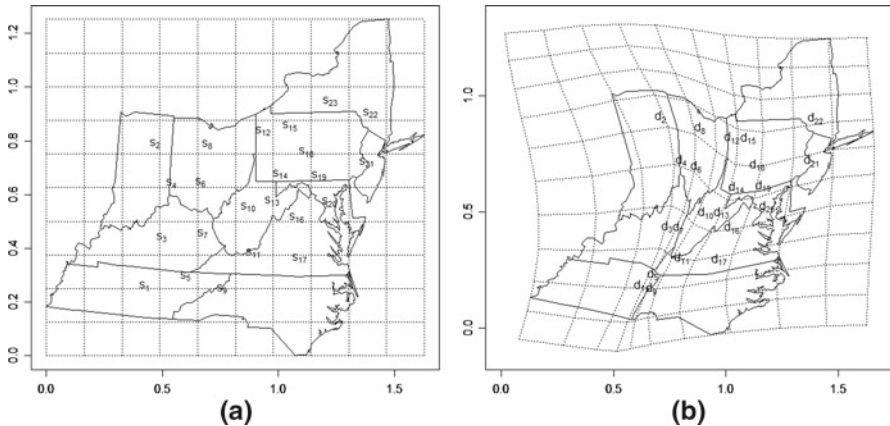




**Fig. 7** Posterior distribution (solid line) and prior distribution (dashed line) for the parameters  $b$ , using the prior distributions  $b \sim G(6 \times 52, 52)$ ,  $\sigma_{d_{ii}}^2 \sim IG(1000 \times 100, 100)$ ,  $i = 1, 2$  (left hand side) and  $\sigma_{d_{ii}}^2 \sim IG(100 \times 0.01 + 1, 0.01)$  (right hand side)



**Fig. 8** Posterior distribution (solid line) and prior distribution (dashed line) for the parameter  $\sigma_{d_{11}}^2$  using the prior distribution  $b \sim G(6 \times 52, 52)$ , and the prior distribution  $\sigma_{d_{ii}}^2 \sim IG(1000 \times 100, 100)$ ,  $i = 1, 2$  (left hand side);  $\sigma_{d_{ii}}^2 \sim IG(100 \times 0.01 + 1, 0.01)$  (right hand side)



**Fig. 9** **a** Monitoring stations in eastern United States; **b** estimated spatial deformation under  $b \sim G(6 \times 52, 52)$ ,  $\sigma_{d_{ii}}^2 \sim IG(1000 \times 100, 100)$ ,  $i = 1, 2$

We proceed with the analysis of the fitted model using the first prior specification for the parameters  $\sigma_{d_{ii}}^2$ ,  $i = 1, 2$ . This application clearly shows how geographic characteristics of the region of study can influence the correlation structure of the spatial

process. Figure 9 shows the spatial deformation obtained in this application. An interesting characteristic of the obtained deformation is that it approximates monitoring stations which are located in the Appalachian mountains, producing the visual effect of a shrinkage in the middle of the map, in a diagonal direction.

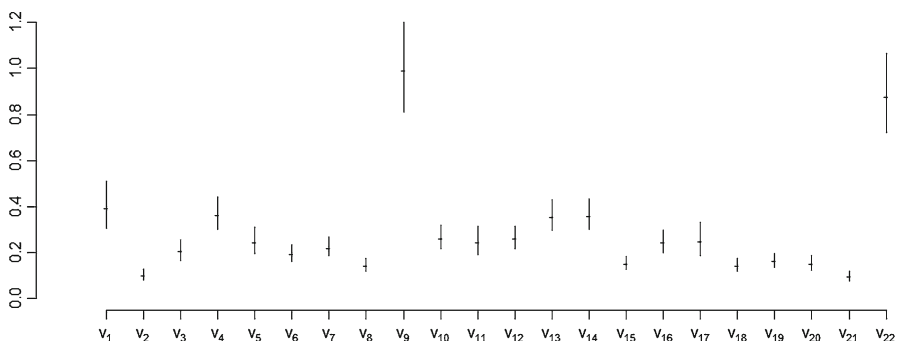
Table 4 presents the estimated posterior mean and 95 % credibility intervals for the parameters  $\sigma_{d11}^2$ ,  $\sigma_{d22}^2$ ,  $b$ ,  $\tau^2$ , and the elements of the main diagonal of matrix  $\Psi$ , denoted by  $\Psi_{ii}$ ,  $i = 1, \dots, 5$ . Figure 10 illustrates the 95 % credibility intervals and the estimated mean of the variances  $v_i$ ,  $i = 1, \dots, 22$ , showing higher variability in the monitoring stations 9 and 22.

Figure 11 presents the estimated mean and 95 % credibility intervals for the dynamic regression coefficients. It can be seen that the coefficients of the intercept and of the seasonal component vary significantly over time, justifying the use of the dynamic specification for these components.

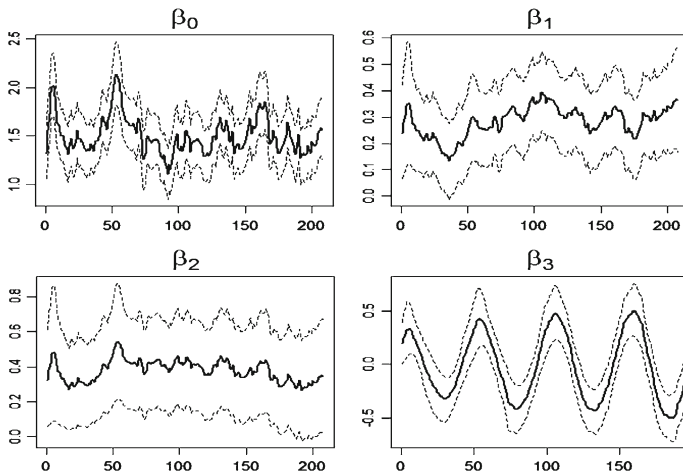
Figure 12 presents the observed time series for some monitoring stations, with their estimated values and credibility intervals. It can be seen that the fitted model leads to good predictions in general.

**Table 4** Posterior mean and credibility intervals of the parameters  $\sigma_{dii}^2$ ,  $i = 1, 2$ ,  $b$ ,  $\tau^2$ , and the elements of the main diagonal of matrix  $\Psi$

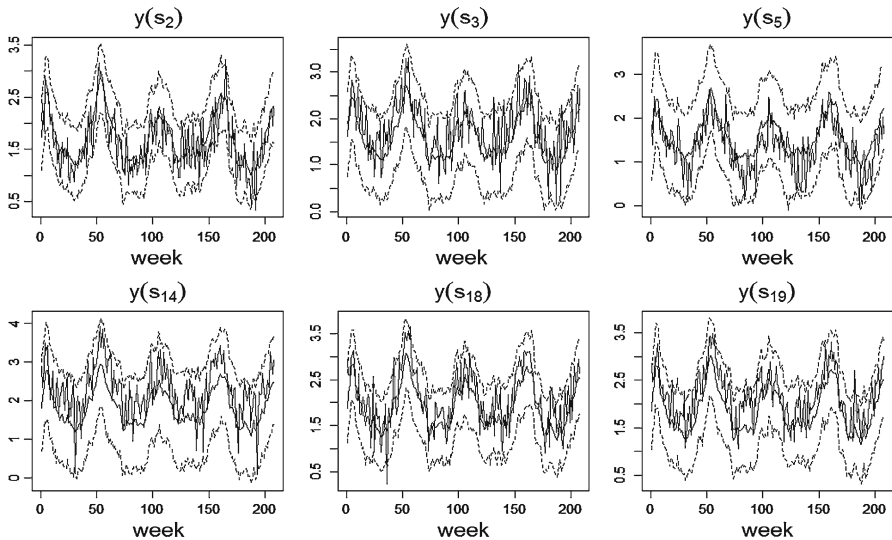
Parameter	Posterior mean	2.5 %	97.5 %
$\sigma_{d11}^2$	0.00101	0.00090	0.00101
$\sigma_{d22}^2$	0.00100	0.00099	0.00101
$b$	6.11	5.61	6.64
$\tau^2$	0.20	0.24	0.28
$\Psi_{11}$	0.0153	0.0092	0.0239
$\Psi_{22}$	0.0011	0.0005	0.0023
$\Psi_{33}$	0.0017	0.0005	0.0039
$\Psi_{44}$	0.000463	0.000054	0.001386
$\Psi_{55}$	0.000205	0.000045	0.000663



**Fig. 10** Estimated mean and 95 % credibility intervals for the variances  $v_i$ ,  $i = 1, \dots, 22$



**Fig. 11** Estimated mean (solid line) and 95 % credibility intervals (dashed line) for  $\beta_i$

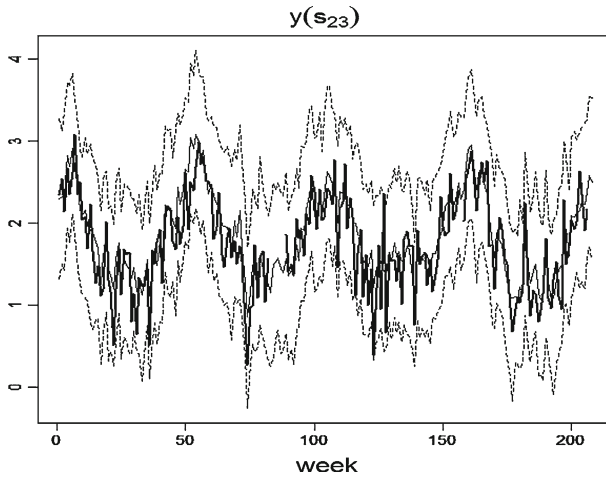


**Fig. 12** True values (in bold), estimated mean (solid line) and 95 % credibility intervals (dashed line) of the response variable  $\log(\text{SO}_2)$  for the monitoring stations in locations  $s_2, s_3, s_5, s_{14}, s_{18}$  and  $s_{19}$

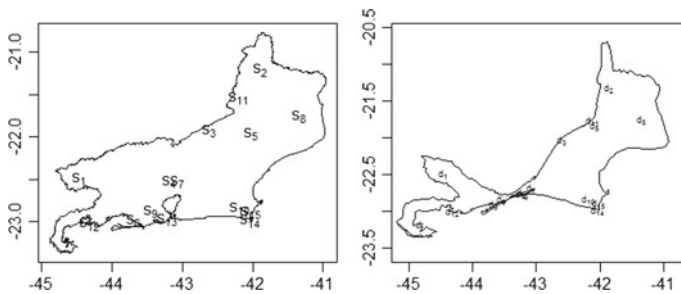
Figure 13 shows the observed time series at station  $s_{23}$ , which was not used to fit the model, and the estimated values and credibility intervals obtained for that station. The figure shows a good fit of the model to the observed data, which indicates that the proposed model is able to produce good predictions for ungauged locations.

## 5.2 Minimum temperature in the State of Rio de Janeiro

In this example we worked with the minimum monthly temperature (measured in Celsius degrees) observed at  $n = 15$  monitoring stations in locations  $\{s_1, \dots, s_{15}\}$  in



**Fig. 13** True values (in *bold*), estimated mean (solid line) and 95 % credibility intervals (dashed line) of the response variable  $\log(\text{SO}_2)$  for the monitoring station  $s_{23}$ , which was not used to fit the model



**Fig. 14** **a** Monitoring Stations of temperature in the State of Rio de Janeiro, **b** estimated spatial deformation in the geographic region of study ( $G$ -space) to  $D$ -space

the State of Rio de Janeiro (see Fig. 14). This data-set was made available by Instituto Nacional de Meteorologia (INMET) and the observations correspond to  $n$  time series observed between January 1961 and December 2000.

Latitude ( $X_1$ ), longitude ( $X_2$ ) and altitude ( $X_3$ ) of the monitoring stations were used as explanatory variables. Also, an exploratory analysis showed that the size of the early seasonal cycle varies with the altitude. Taking that into account, the monitoring stations were divided into five groups. Variables  $\delta_i$ ,  $i = 1, \dots, 15$ , indicating to each group each station belongs, were created as below

$$\delta_i = \begin{cases} (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0), & \text{if } x_{3i} \in (0, 100], \\ (0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0), & \text{if } x_{3i} \in (100, 350], \\ (0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0), & \text{if } x_{3i} \in (350, 450], \\ (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0), & \text{if } x_{3i} \in (450, 650], \\ (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1), & \text{if } x_{3i} > 650, \end{cases}$$

where  $\delta = (\delta_1, \dots, \delta_{15})'$ , and  $\mathbf{x}_{ki} = (x_{k1}, \dots, x_{k,15})'$ ,  $k = 1, 2, 3$ , are vectors corresponding to the latitude, longitude and altitude at the monitoring stations.

The proposed model is specified as in (1)–(7). The mean of the process was modeled assuming a second order dynamic linear structure, which is a structure that incorporates an extra time-varying parameter representing a growth in the level of the series. We define  $\mathbf{F}'_t = (\mathbf{1}_{15}, \mathbf{0}_{15}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \delta)$ , and the coefficients associated to  $\mathbf{F}_t$  are allowed to vary in time, so that the effects of the covariates may change over time. This evolution is defined through  $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_1 & \mathbf{0}_{2 \times 3} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{3 \times 2} & \mathbf{I}_3 & \mathbf{0}_{3 \times 2} \\ \mathbf{0}_{10 \times 2} & \mathbf{0}_{10 \times 3} & \mathbf{1}_5 \otimes \mathbf{G}_2 \end{pmatrix}$ , with  $\mathbf{G}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  and  $\mathbf{G}_2 = \begin{pmatrix} \cos \frac{\pi}{6} & \sin \frac{\pi}{6} \\ -\sin \frac{\pi}{6} & \cos \frac{\pi}{6} \end{pmatrix}$ .

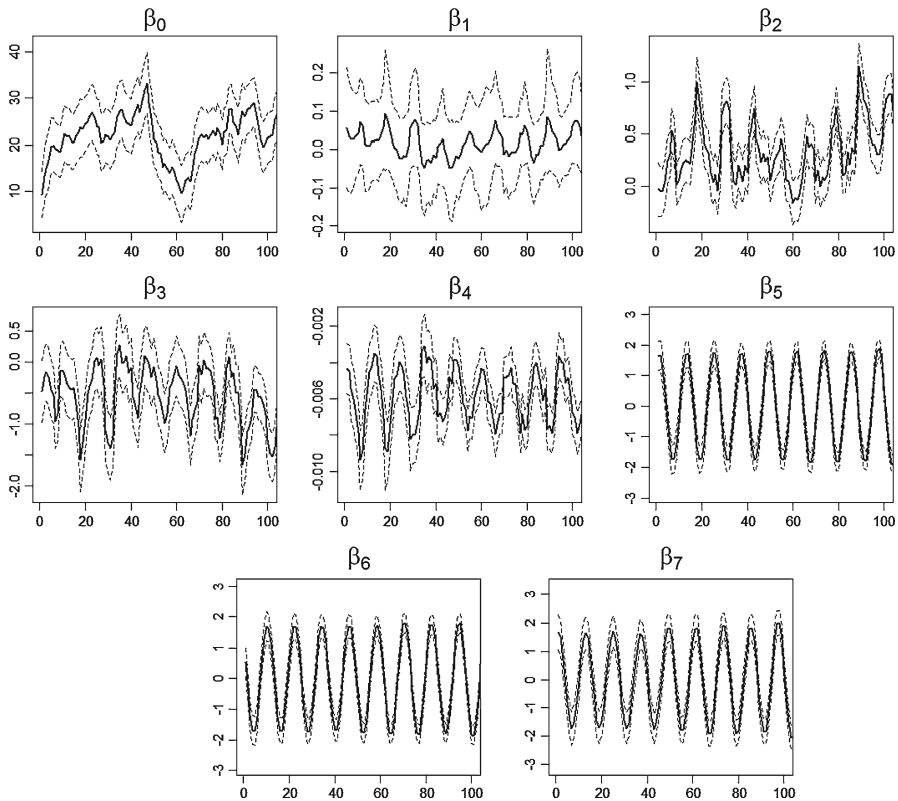
The model is completed with the specification of prior distributions for the parameters  $\mathbf{v}$  and  $b$ . The use of non informative priors for the parameters  $\sigma_{di}^2$ ,  $i = 1, 2$  led again to convergence problems in the MCMC chains. A number of informative prior specifications were tested to correct the anisotropy in the model. The prior distribution for the parameter  $\sigma_{di}^2$ ,  $i = 1, 2$ , was chosen to be a  $IG(220 \times 1000, 1000)$  based on these preliminary trials.

Figure 14 shows the State of Rio de Janeiro with monitoring stations in the original geographic space and also the estimated spatial deformation. The spatial deformation shows a compression in the mountain region. Figure 15 shows the estimated mean and the 95 % credibility intervals of  $\beta_i$ ,  $i = 1, \dots, 5$ , showing a significant variation of most regression coefficients in time, and justifying the use of dynamic models. Figure 16 shows the estimated mean and the 95 % credibility intervals of the variances  $v_i$ ,  $i = 1, \dots, 22$ , and it shows that the monitoring station at location  $s_4$  shows greater variability than all the others.

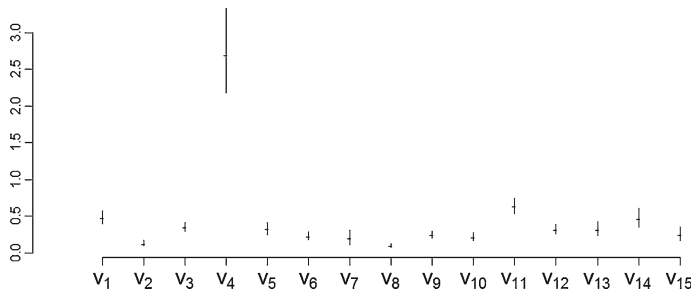
## 6 Final remarks

In this paper we proposed a model for nonstationary space time observations, which models the mean of the process based on the approach of State Space models (West and Harrison 1997) and corrects anisotropy through spatial deformation (Sampson and Guttorp 1992) under the Bayesian point of view (Schmidt and O'Hagan 2003). In this model all of the parameters, including the parameters of the mean, are estimated via the Bayes paradigm. This new approach allows, in a simple way, to model the anisotropic covariance structure, perform predictions of the process for non-monitored locations and treat missing observations. It also shows an advantage over other existing models in the literature, specially on the grounds of interpolation performance, as exemplified through simulation studies.

One disadvantage of the proposed approach is that when a strong spatial deformation is needed to correct the anisotropy, informative priors must be specified for the parameters  $\sigma_{di}^2$ ,  $i = 1, 2$ . Gelman (2006) recommends using prior distributions in the semi-Cauchy family when the scale parameter  $\sigma$  presents values close to zero, instead of using the noninformative priors  $G(\zeta, \zeta)$  with  $\zeta \rightarrow 0$ , as the estimation of the parameters of interest can be sensitive to the choice of  $\zeta$ . Future work could include a



**Fig. 15** Estimated mean (solid line) and 95 % credibility intervals (dashed line) for the coefficients  $\beta_0, \dots, \beta_7$



**Fig. 16** Estimated mean and 95 % credibility intervals of the variances  $v_i, i = 1, \dots, 15$

sensitivity study of the proposed model, using prior specifications in the semi-Cauchy family, in cases when strong deformations are needed to correct the anisotropy.

Gamerman et al. (2007) suggest that the spatial-temporal covariance structure could be affected by space and time. They also suggest that a natural candidate to introduce this dependence is given by dynamic Gaussian processes. Another future work is to propose a spatial-temporal approach for nonstationary Gaussian processes, where the mean of the process is modeled via dynamic models (West and Harrison 1997) and the

spatial covariance structure via a dynamic generalization of the spatial deformation. Partial results of this new methodology can be seen in [Castro \(2010\)](#).

**Acknowledgments** The authors thank the referees and Editor for many useful comments. This paper is based on parts of the PhD project of the first author under the supervision of the second author. The second author has benefitted from research grants awarded by CNPq and FAPERJ, and the third author has benefited from research grants awarded by CNPq.

## References

- Berger JO, Oliveira V, Sanso B (2001) Objective Bayesian analysis of spatially correlated data. *J Am Stat Assoc* 96:1361–1374
- Box GEP, Jenkins GM (1976) Time series analysis forecasting and control. Holden-Day, San Francisco
- Bruno F, Cocchi D (2004) Non-separability in spatiotemporal covariance models arisen by non stationarity in time. In: *Atti della XLII Riunione Scientifica SIS, CLEUP*, pp. 157–160
- Bruno F, Guttorp P, Sampson PD, Cocchi D (2008) A simple non-separable, non-stationary spatiotemporal model for ozone. *Environ Ecol Stat* 16:515–529
- Casella G, Berger RL (2001) Statistical inference, 2nd edn. Duxbury Press, Belmont, CA
- Castro F (2010) Dynamic models for spatial deformations (in Portuguese). Unpublished PhD Thesis, Instituto de Matemática, DME-UFRJ, Rio de Janeiro
- Damian D, Sampson PD, Guttorp P (2001) Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics* 12:161–178
- Gamerman D, Lopes HF (2006) Markov Chain Monte Carlo—stochastic simulation for Bayesian inference, 2nd edn. Chapman & Hall/CRC, London
- Gamerman D, Migon HS (1999) Statistical inference: an integrated approach. Arnold, London
- Gamerman D, Salazar E, Reis EA (2007) Dynamic Gaussian process priors, with applications to the analysis of space-time data (with discussion). In: Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM, West M (eds) (Org) Bayesian statistics, vol 8. Oxford Univeristy Press, Oxford, pp 149–174
- Gelman A (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Anal* 1: 515–533
- Harvey AC (1989) Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge
- Hastings W (1970) Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* 57:97–109
- Kendall DG (1984) Shape manifolds, Procrustean metrics and complex projective spaces. *Bull Lond Math Soc* 16:81–121
- Mardia KV, Kent JT, Walder AN (1991) Statistical shape models in image analysis. In: *Proceedings of the 12th symposium on the interface between computing science and statistics*. Interface Foundation, Fairfax Station, pp 449–446
- Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E (1953) Equation of state calculations by fast computing machine. *J Chem Phys* 21:1087–1091
- Robert CR, Casella G (2004) Monte Carlo statistical methods. Springer, New York
- Sampson PD, Guttorp P (1992) Nonparametric estimation of nonstationary spatial covariance structure. *J Am Stat Assoc* 87:108–119
- Sampson PD, Damian D, Guttorp P (2001) Advances in modeling and inference for environmental processes with nonstationary spatial covariance. In: Monestiez P, Aillard D, Froidevaux R (eds) *GeoENV 2000: geostatistics for environmental applications*. Kluwer, Dordrecht pp 17–32
- Schmidt AM, Gelfand A (2003) A Bayesian coregionalization model for multivariate pollutant data. *J Geophys Res* 108(D24). <http://www.agu.org/pubs/crossref/2003/2002JD002905.shtml>
- Schmidt AM, O'Hagan A (2003) Bayesian inference for nonstationary spatial covariance structure via spatial deformations. *J R Stat Soc Ser B* 65:743–758
- Spiegelhalter D, Best N, Carlin B, Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Ser B* 64:583–639
- West M, Harrison J (1997) Bayesian forecasting and dynamic models. 2. Springer, New York

## Author Biographies

**Fidel Ernesto Castro Morales** First degree in Industrial Engineering from Universidad Nacional de Colombia-Sede Medellin (UNAL) in 2003, M.Sc. in Statistics from UNAL in 2004 and Ph.D. in Statistics from UFRJ in 2010. Associate Professor of Statistics at UFRN since 2010. Current research interests include space-time models, spatial statistics, dynamic models and Bayesian inference.

**Dani Gamerman** First degree in Mechanical Engineering from the Military Engineering Institute in 1980. M.Sc. in Statistics from IMPA in 1983. Ph.D. in Statistics from University of Warwick in 1987. Professor of Statistics at UFRJ since 1996. Supervises M. Sc. and Ph.D. students. Invited lecturer at various scientific meetings in Brazil and abroad. Visiting lecturer of a few Universities in Brazil and abroad and colaborador honorífico of Universidade Rey Juan Carlos in Madrid. Author of the books Monte Carlo Markov Chain: Stochastic Simulation for Bayesian Inference, published by Chapman and Hall in 1997 (1st edition) and 2006 (2nd edition, with Hedibert F. Lopes) and Statistical Inference: an Integrated Approach (with Helio S. Migon), published by Arnold in 1999. Papers published in many statistical journals, including Journal of the Royal Statistical Society, Series B and Biometrika. Current research interests include dynamic models, item response theory, spatial statistics, survival analysis, stochastic simulation, econometrics and Bayesian inference.

**Marina Silva Paez** First degree in Statistics from Universidade Federal do Rio de Janeiro (UFRJ) in 1998, M.Sc. in Statistics from UFRJ in 2000 and Ph.D. in Statistics from UFRJ in 2004. Was a research student at Lancaster University from September 2003 to August 2004. Associate Professor of Statistics at UFRJ since 2005. Supervises M.Sc. and Ph.D. students. Have published papers in statistical journals, including Environmetrics and Environmental and Ecological Statistics, mainly on the subject of space-time Bayesian models. Invited speaker at many scientific meeting in Brazil. Current research interests include space-time models, spatial statistics, dynamic models, item response theory, and Bayesian inference.